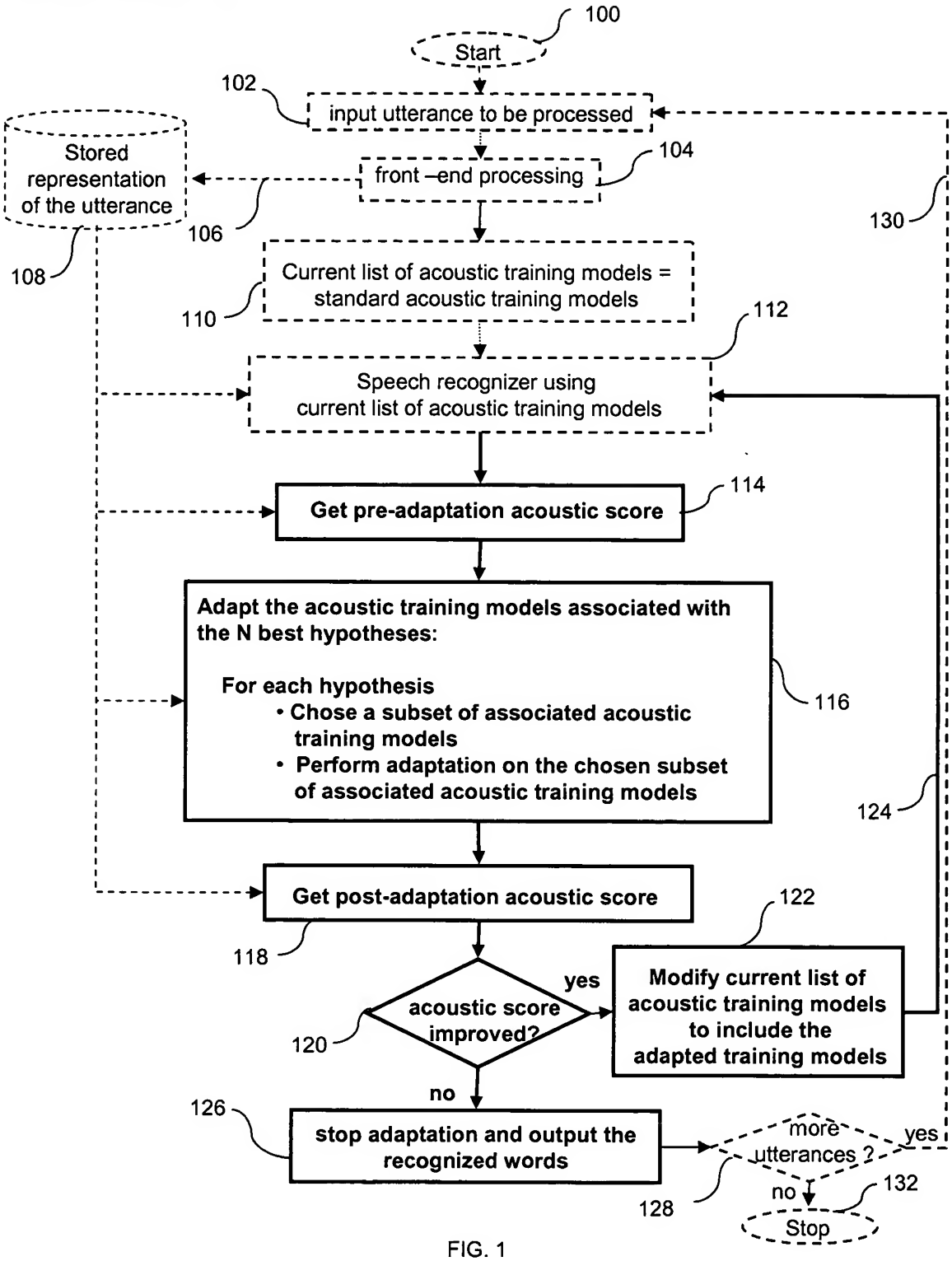

100
Start
102
input utterance to be processed
Stored representation of the utterance
104
front –end processing
106
130
108
Current list of acoustic training models = standard acoustic training models
110
112
Speech recognizer using current list of acoustic training models
114
Get pre-adaptation acoustic score
Adapt the acoustic training models associated with the N best hypotheses:
For each hypothesis
• Chose a subset of associated acoustic training models
• Perform adaptation on the chosen subset of associated acoustic training models
116
124
Get post-adaptation acoustic score
122
118
yes
acoustic score improved?
120
Modify current list of acoustic training models to include the adapted training models
no
126
stop adaptation and output the recognized words
more utterances ?
yes
no
128
132
Stop

FIG. 1

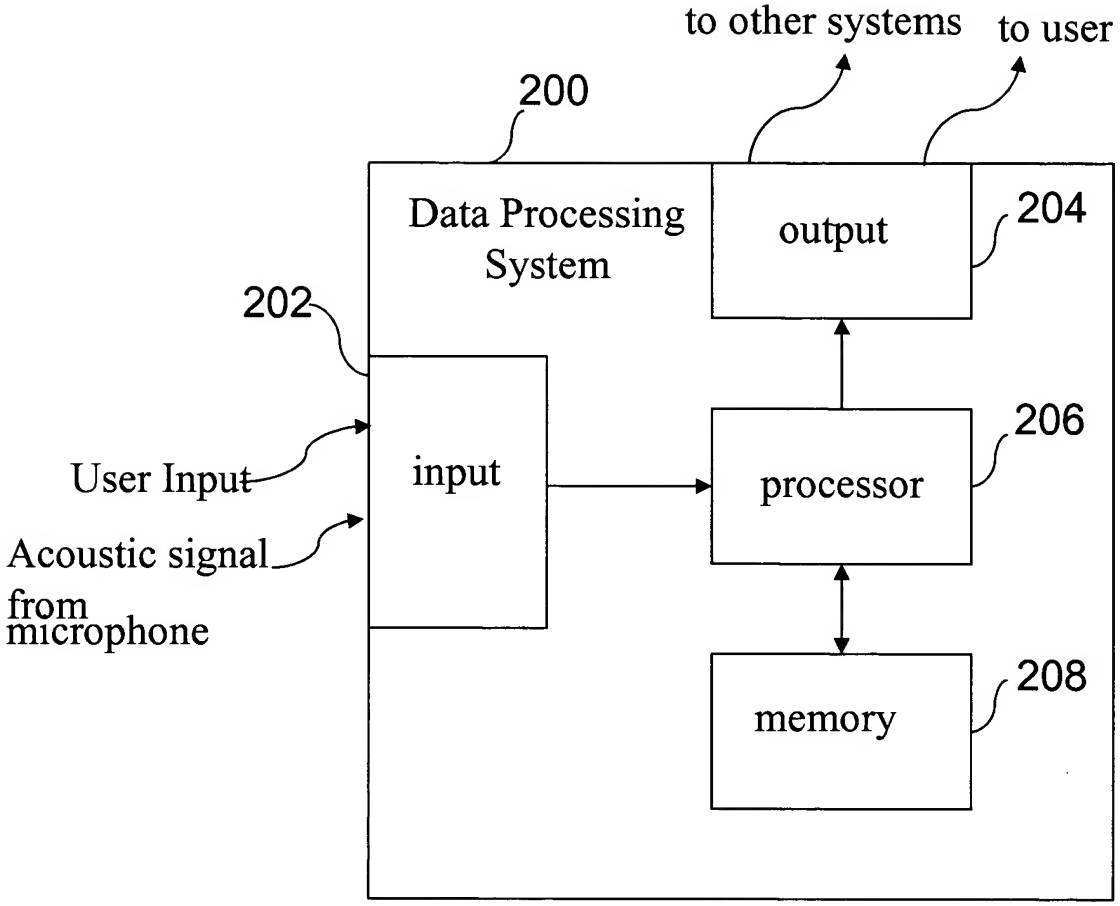

to other systems
to user
200
Data Processing System
output
204
202
input
processor
206
User Input
Acoustic signal from microphone
memory
208

FIG. 2

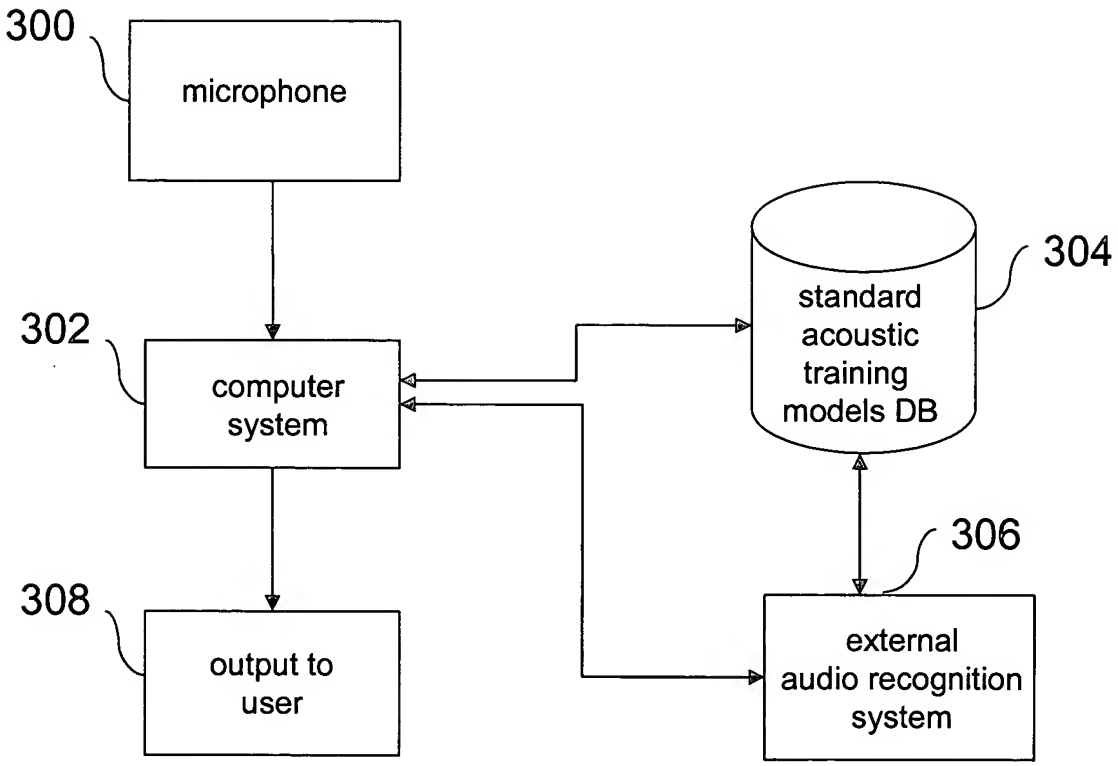

300
microphone
302
computer system
304
standard acoustic training models DB
306
external audio recognition system
308
output to user

FIG. 3

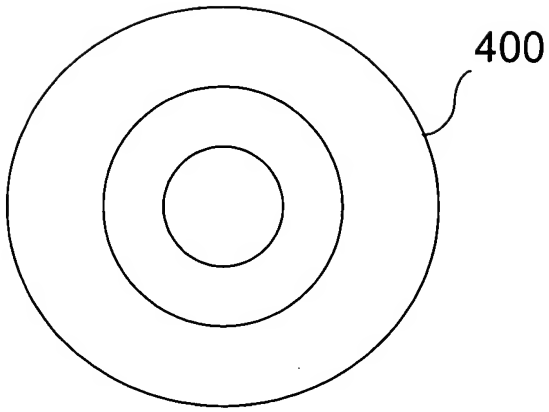


FIG. 4

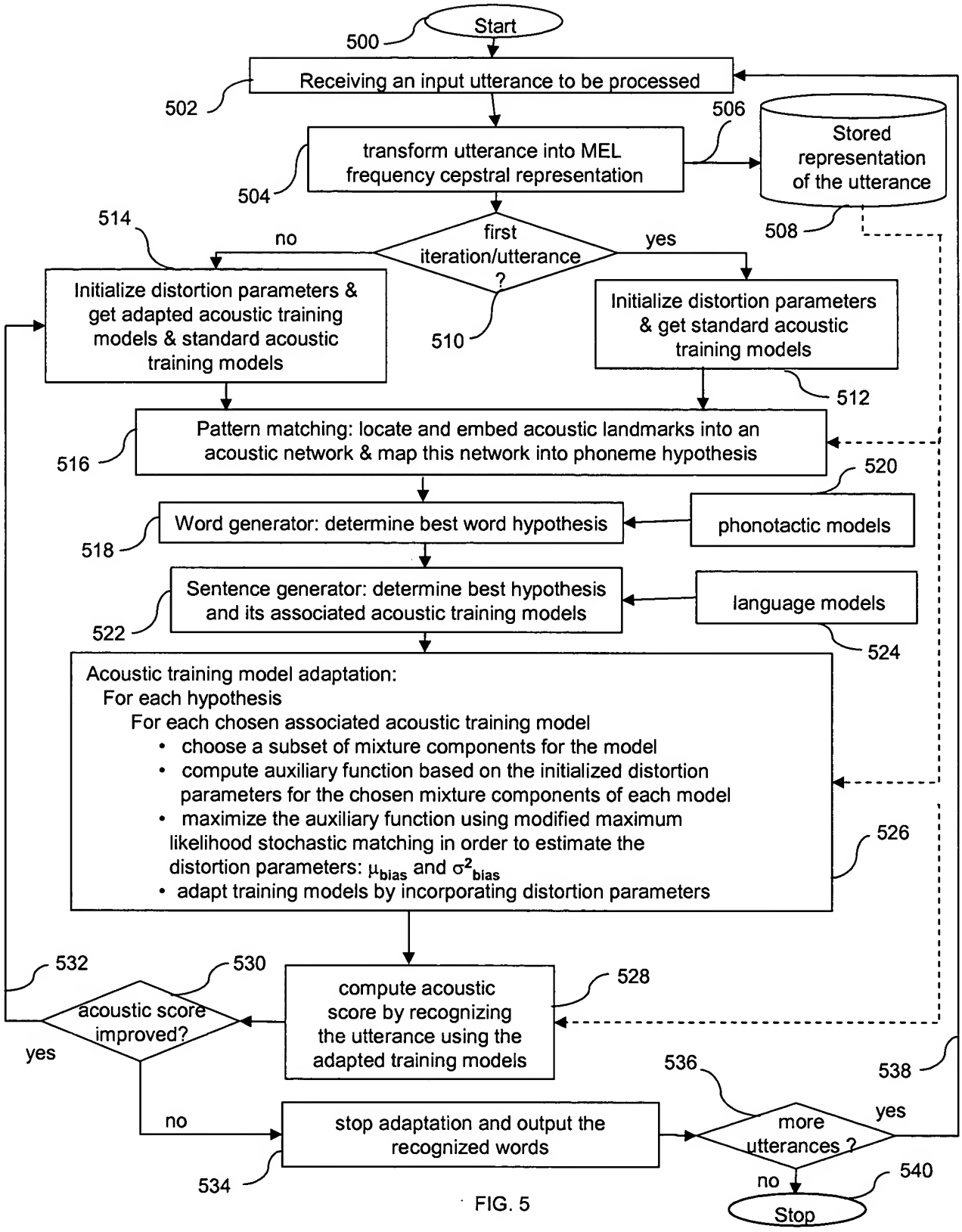

Start
500
Receiving an input utterance to be processed
502
transform utterance into MEL frequency cepstral representation
504
506
Stored representation of the utterance
508
first iteration/utterance ?
510
no
yes
514
Initialize distortion parameters & get adapted acoustic training models & standard acoustic training models
Initialize distortion parameters & get standard acoustic training models
512
Pattern matching: locate and embed acoustic landmarks into an acoustic network & map this network into phoneme hypothesis
516
520
phonotactic models
Word generator: determine best word hypothesis
518
language models
Sentence generator: determine best hypothesis and its associated acoustic training models
522
524
Acoustic training model adaptation:
For each hypothesis
For each chosen associated acoustic training model
• choose a subset of mixture components for the model
• compute auxiliary function based on the initialized distortion parameters for the chosen mixture components of each model
• maximize the auxiliary function using modified maximum likelihood stochastic matching in order to estimate the distortion parameters: μ_{bias} and σ^2_{bias}
• adapt training models by incorporating distortion parameters
526
532
530
acoustic score improved?
yes
compute acoustic score by recognizing the utterance using the adapted training models
528
536
538
no
stop adaptation and output the recognized words
534
more utterances ?
yes
no
540
Stop

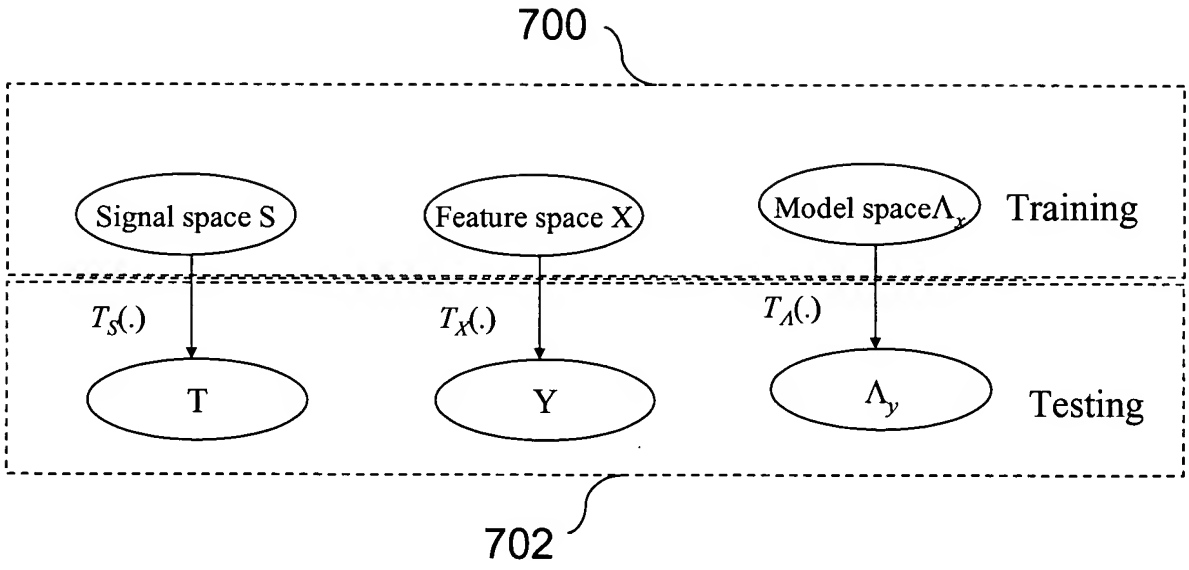
FIG. 5

600

SNR	clean	12dB	6 dB	0dB
% Accurate	91.9	82.3	68.5	45.2

Recognition accuracy (word level) of noisy speech signals

FIG. 6



Conceptual description of adaptation

FIG. 7